Project 1: Credit Risk Assessment:

Problem Statement: Develop a machine learning model to assess the credit risk of individuals based on their financial and personal information.

Daily Tasks:

Day 1:

Introduction to credit risk assessment and its importance in banking and finance.

Overview of classification algorithms suitable for this task.

Day 2:

Data set preparation: Collect a dataset containing individual credit profiles and risk labels.

Perform data preprocessing, handling missing values, categorical features, and feature scaling.

Day 3:

Conduct exploratory data analysis (EDA) to understand the dataset and identify relevant features.

Preprocess the data further, such as encoding categorical variables or handling imbalanced classes.

Day 4:

Implement a baseline classification model, such as logistic regression or a random forest classifier.

Explain evaluation metrics for credit risk assessment, such as accuracy, precision, recall, and F1 score.

Day 5:

Improve the model's performance through techniques like feature engineering or ensemble models.

Discuss strategies for handling class imbalance, such as oversampling or under sampling.

Day 6:

Evaluate the model's performance on a test set and discuss the results.

Identify limitations and suggest potential improvements.

Prepare a presentation summarizing the findings and provide feedback.

Evaluation criteria: Model accuracy, precision, recall, F1 score, and presentation quality.

Dataset Format: Ensure the dataset contains individual credit profiles with relevant columns such as age, income, employment status, loan amount, loan duration, credit score, and risk labels (e.g., high risk, low risk).

Data Types: Numeric columns for age, income, loan amount, and credit score. Categorical columns for employment status and risk labels.

Project 2: Spam Email Classification

Problem Statement: Develop a machine learning model to classify emails as spam or non-spam based on their content and metadata.

Daily Tasks:

Day 1:

Introduction to spam email classification and its impact on email filtering.

Overview of classification algorithms suitable for this task.

Day 2:

Data set preparation: Collect a dataset containing email samples and corresponding labels.

Perform text data preprocessing, such as removing stop words, stemming, or vectorization.

Day 3:

Conduct exploratory data analysis (EDA) to gain insights into the dataset and identify important features.

Perform feature extraction or selection, such as using TF-IDF or word embeddings.

Day 4:

Implement a baseline classification model, such as Naive Bayes or a support vector machine (SVM) classifier.

Explain evaluation metrics for spam email classification, such as accuracy, precision, recall, and F1 score.

Day 5:

Improve the model's performance through techniques like feature engineering or ensemble models.

Discuss strategies for handling class imbalance, if present, such as adjusting classification thresholds.

Day 6:

Evaluate the model's performance on a test set and discuss the results.

Identify limitations and suggest potential enhancements.

Prepare a presentation summarizing the findings and provide feedback.

Evaluation criteria: Model accuracy, precision, recall, F1 score, and presentation quality.

Dataset Format: Collect a dataset with email samples and corresponding labels (spam or non-spam). The dataset should include email content and metadata such as sender, subject, and timestamp.

Data Types: Text columns for email content, categorical column for labels (spam or non-spam), and string columns for sender, subject, and timestamp.

Project 3: Customer Segmentation

Problem Statement: Cluster customers into distinct groups based on their demographics, behaviors, or purchase patterns.

Daily Tasks:

Day 1:

Introduction to customer segmentation and its applications in marketing and business strategies.

Overview of clustering algorithms suitable for this task.

Day 2:

Data set preparation: Collect a dataset containing customer information and attributes.

Perform data preprocessing, such as handling missing values, categorical features, and feature scaling.

Day 3:

Conduct exploratory data analysis (EDA) to understand the dataset and identify relevant features.

Preprocess the data further, such as encoding categorical variables or scaling numerical features.

Day 4:

Implement a baseline clustering model, such as K-means or hierarchical clustering.

Explain evaluation metrics for clustering, such as silhouette score or within-cluster sum of squares.

Day 5:

Improve the model's performance through techniques like feature engineering or alternative clustering algorithms.

Discuss strategies for selecting the optimal number of clusters, such as the elbow method or silhouette analysis.

Day 6:

Evaluate the model's performance using appropriate clustering metrics and interpret the results.

Identify limitations and suggest potential improvements.

Prepare a presentation summarizing the findings and provide feedback.

Evaluation criteria: Clustering metrics, interpretation of results, presentation quality.

Dataset Format: Obtain a dataset with customer information and attributes, including demographic variables (age, gender, location), behavioral variables (purchase history, website visits), and customer value metrics (total spend, frequency).

Data Types: Numeric columns for customer value metrics, categorical columns for demographic variables, and potentially numeric or categorical columns for behavioral variables.

Project 4: Stock Price Prediction

Problem Statement: Develop a machine learning model to predict stock prices based on historical market data.

Daily Tasks:

Day 1:

Introduction to stock price prediction and its relevance in financial analysis and investment strategies.

Overview of regression algorithms suitable for this task.

Day 2:

Data set preparation: Collect a dataset containing historical stock market data and corresponding prices.

Perform data preprocessing, such as handling missing values, feature scaling, and feature engineering.

Day 3:

Conduct exploratory data analysis (EDA) to understand the dataset and identify relevant features.

Preprocess the data further, such as handling outliers or transforming variables.

Day 4:

Implement a baseline regression model, such as linear regression or support vector regression (SVR).

Explain evaluation metrics for stock price prediction, such as mean squared error (MSE) or root mean squared error (RMSE).

Day 5:

Improve the model's performance through techniques like feature engineering, time series analysis, or ensemble models.

Discuss strategies for handling non-stationary data or incorporating external factors into the model.

Day 6:

Evaluate the model's performance using appropriate regression metrics and interpret the results.

Identify limitations and suggest potential enhancements.

Prepare a presentation summarizing the findings and provide feedback.

Evaluation criteria: Model performance metrics (e.g., MSE, RMSE), interpretation of results, presentation quality.

Dataset Format: Collect a dataset containing historical stock market data, including columns such as date, opening price, closing price, highest price, lowest price, and trading volume.

Data Types: Date column for the date, numeric columns for prices and volume.

Project 5: Loan Approval Prediction

Problem Statement: Develop a machine learning model to predict the likelihood of loan approval based on applicants' information.

Daily Tasks:

Day 1:

Introduction to loan approval prediction and its significance in the banking and lending industry.

Overview of classification algorithms suitable for this task.

Day 2:

Data set preparation: Collect a dataset containing applicant information and loan approval labels.

Perform data preprocessing, such as handling missing values, categorical features, and feature scaling.

Day 3:

Conduct exploratory data analysis (EDA) to understand the dataset and identify relevant features.

Preprocess the data further, such as encoding categorical variables or handling imbalanced classes.

Day 4:

Implement a baseline classification model, such as logistic regression or a decision tree classifier.

Explain evaluation metrics for loan approval prediction, such as accuracy, precision, recall, and F1 score.

Day 5:

Improve the model's performance through techniques like feature engineering or ensemble models.

Discuss strategies for handling class imbalance, such as oversampling or under sampling.

Day 6:

Evaluate the model's performance on a test set and discuss the results.

Identify limitations and suggest potential improvements.

Prepare a presentation summarizing the findings and provide feedback.

Evaluation criteria: Model accuracy, precision, recall, F1 score, and presentation quality.

Dataset Format: Gather a dataset with applicant information, including columns such as age, income, employment status, loan amount, loan purpose, and loan approval status.

Data Types: Numeric columns for age, income, loan amount, and categorical columns for employment status, loan purpose, and loan approval status.

Project 6: Customer Lifetime Value Prediction

Problem Statement: Develop a machine learning model to predict the potential value of a customer over their lifetime with a company.

Daily Tasks:

Day 1:

Introduction to customer lifetime value (CLV) and its importance in customer relationship management.

Overview of regression algorithms suitable for this task.

Day 2:

Data set preparation: Collect a dataset containing customer transaction history and CLV values.

Perform data preprocessing, such as handling missing values, categorical features, and feature scaling.

Day 3:

Conduct exploratory data analysis (EDA) to understand the dataset and identify relevant features.

Preprocess the data further, such as handling outliers or transforming variables.

Day 4:

Implement a baseline regression model, such as linear regression or a decision tree-based algorithm.

Explain evaluation metrics for CLV prediction, such as mean squared error (MSE) or root mean squared error (RMSE).

Day 5:

Improve the model's performance through techniques like feature engineering, time series analysis, or ensemble models.

Discuss strategies for handling non-linear relationships or incorporating customer segmentation into the model.

Day 6:

Evaluate the model's performance using appropriate regression metrics and interpret the results.

Identify limitations and suggest potential enhancements.

Prepare a presentation summarizing the findings and provide feedback.

Evaluation criteria: Model performance metrics (e.g., MSE, RMSE), interpretation of results, presentation quality.

Dataset Format: Obtain a dataset with customer transaction history, including columns such as customer ID, transaction date, purchase amount, and customer lifetime value (CLV).

Data Types: Numeric columns for purchase amount and CLV, categorical column for customer ID, and date column for transaction date.